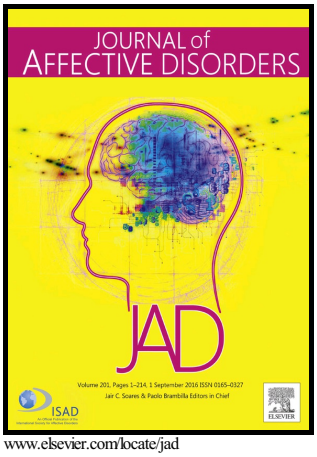
DOI:
10.1016/j.jad.2017.11.037

Document Version
Peer reviewed version



www.elsevier.com/locate/jad

Evaluation of the Hospital Anxiety and Depression Scale (HADS) in Screening Stroke Patients for Symptoms: Item Response Theory (IRT) Analysis

Salma A Ayis, Luis Ayerbe, Mark Ashworth, Charles DA Wolfe

PII: S0165-0327(17)31132-1
DOI: http://dx.doi.org/10.1016/j.jad.2017.11.037
Reference: JAD9351

To appear in: *Journal of Affective Disorders*

Received date: 4 June 2017
Revised date: 10 September 2017
Accepted date: 11 November 2017

Cite this article as: Salma A Ayis, Luis Ayerbe, Mark Ashworth and Charles DA Wolfe, Evaluation of the Hospital Anxiety and Depression Scale (HADS) in Screening Stroke Patients for Symptoms: Item Response Theory (IRT) Analysis *Journal of Affective Disorders,* http://dx.doi.org/10.1016/j.jad.2017.11.037

Evaluation of the Hospital Anxiety and Depression Scale (HADS) in Screening Stroke Patients for Symptoms: Item Response Theory (IRT) Analysis

Salma A Ayis PhD[1,2], Luis Ayerbe PhD[3], Mark Ashworth DM[1], Charles DA Wolfe FFPH[1,2]

[1]School of Population Sciences and Health Services Research; Faculty of Life Sciences and Medicine King's College London, UK.

[2]National Institute for Health Research (NIHR) Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust, London UK, and the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) South London at King's College Hospital NHS Foundation Trust.

[3]Blizard Institute. Centre for Primary Care and Public Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK.

Corresponding author's address:

Salma Ayis (PhD), Senior Lecturer in Medical Statistics

Department of Primary Care and Public Health Sciences, King's College London, 4th Floor Addison House, Guy's Campus, London SE1 1UL, UK

Tel: + 44 (0) 207 848 8222; Fax: + 44 (0) 207 848 6620

Abstract

Background:

Variations have been reported in the number of underlying constructs and choice of thresholds that determine caseness of anxiety and /or depression using the Hospital Anxiety and Depression scale (HADS). This study examined the properties of each item of HADS as perceived by stroke patients, and assessed the information these items convey about anxiety and depression between 3 months to 5 years after stroke.

Methods:

The study included 1443 stroke patients from the South London Stroke Register (SLSR). The dimensionality of HADS was examined using factor analysis methods, and items' properties up to 5 years after stroke were tested using Item Response Theory (IRT) methods, including graded response models (GRMs).

Results:

The presence of two dimensions of HADS (anxiety and depression) for stroke patients was confirmed. Items that accurately inferred about the severity of anxiety and depression, and offered good discrimination of caseness were identified as "I can laugh and see the funny side of things" (Q4) and "I get sudden feelings of panic" (Q13), discrimination 2.44 (se=.26), and 3.34 (se=0.35), respectively. Items that shared properties, hence replicate inference were: "I get a sort of frightened feeling as if something awful is about to happen" (Q3) and "Worrying thoughts go through my mind" (Q9). Item properties were maintained over time.

Limitations: approximately 20% of patients were lost to follow up.

Conclusion:

A more concise selection of items based on their properties, would provide a precise approach for screening patients and for an optimal allocation of patients into clinical trials.

Keywords: Anxiety; Depression; Stroke; Hospital Anxiety and Depression Scale (HADS); Item Response Theory (IRT); Graded Response Models (GRMs)

Background

The prevalence rates for depression and anxiety in stroke patients vary considerably; a range of 11% to 61% for depression, and 14% to 28% for anxiety, has been reported. (Barker-Collo, 2007; Hackett & Anderson, 2005; Hackett & Pickles, 2014; Leppavuori, Pohjasvaara, Vataja, Kaste, & Erkinjuntti, 2003; Townend et al., 2007) Methods of assessment, scales

used, heterogeneity of study populations, and choices of cut-off points that determine caseness, were among factors that contributed to differences in these estimates.

Systematic reviews have estimated the prevalence of depression at approximately 30% throughout the first 10 years after stroke. (Ayerbe, Ayis, Wolfe, & Rudd, 2013) Anxiety disorders on the other hand were thought to affect between 20% – 25% of stroke patients at any time after stroke. (Campbell Burton et al., 2013) Both conditions negatively impact on physical function, delay recovery, and depression is associated with high mortality.(Naess, Waje-Andreassen, Thomassen, Nyland, & Myhr, 2006)

The hospital anxiety and depression scale (HADS) was devised to screen for emotional disorders in medical practice. (Zigmond & Snaith, 1983) The scale has been validated in a variety of settings for assessing the severity of depression and anxiety in both primary and secondary care settings. (Bjelland, Dahl, Haug, & Neckelmann, 2002; Snaith, 2003) It has been used and validated in patients with stroke and has shown good performance both self-administered and interviewer administered. (Sagen et al., 2009; Turner et al., 2012); (Aben, Verhey, Lousberg, Lodder, & Honig, 2002)

The number and contents of underlying constructs of the scale have been reported to vary widely across different studies. These have included a single construct (Razavi, Delvaux, Farvacques, & Robaye, 1990) two constructs, (Matsudaira et al., 2009; Mykletun, Stordal, & Dahl, 2001) three (Caci et al., 2003; Dawkins, Cloherty, Gracey, & Evans, 2006) and four constructs.(Andersson, 1993) Single constructs were generally interpreted as assessing distress whereas two constructs were seen as assessing anxiety and depression constrcuts seperately, in consistence with what the scale was intended to measure. Studies suggesting three or four factors have identified additional constructs that are highly correlated with

depression and anxiety, for example agitation (Barth & Martin, 2005) and restlessness.(Caci et al., 2003)

Similar variations have been reported for thresholds that identify caseness, including a range from ≥ 3 to ≥ 11 for anxiety (HADS-A), and from ≥ 4 to ≥ 11 for depression (HADS-D). (Bjelland et al., 2002; Sagen et al., 2009; Terol-Cantero, Cabrera-Perona, & Martín-Aragón, 2015) These variations raise several questions about what HADS actually measures in patients with different health conditions and suggests the need for better understanding of the psychometric properties of the scale, how items are perceived by different groups of respondents, and the impact of these differences on summary scores and thresholds.

Item response theory (IRT) methods were developed to construct and evaluate psychometric instruments. The methods were originally devised by teams of psychologists and educationalists in order to design and evaluate examination questions and assessment tools. (Baker, 1985; Hambleton, Swaminathan, & Rogers, 1991) More recently, these methods have been used to improve several health outcome measures. (Bagby, Ryder, Schuller, & Marshall, 2004; Chakravarty, Bjorner, & Fries, 2007; Hays, Morales, & Reise, 2000; Olino et al., 2013).

IRT models are members of the broader class of latent trait models that estimate underlying unobservable traits using observed variables. The models relate characteristics of items (difficulty and discrimination) and characteristics of individuals (intelligence, distress, or physical function, for example) to the probability of selecting various options of an item in a scale.

The aims of this study are firstly, to assess the psychometric properties of the scale as perceived by stroke patients, and to investigate how each of its items relate to the underlying levels of anxiety and depression; secondly, to test the stability of the properties over five

years following a stroke and thirdly, to highlight the clinical implications of item properties on the determination of caseness.

Methods

Patients:

Participants were recruited from the South London Stroke Register (SLSR), a prospective population-based cohort study.(Heuschmann, Grieve, Toschke, Rudd, & Wolfe, 2008) Recruitment took place between, 1998 and 2013, and patients were followed up to June 2014. From 3,942 registered patients, 892 (22.6%) died, 1236 (31.3%) were lost to follow up. From the remaining 1815 patients, HADS questionnaire was completed by 1443 (79.5%) at 3 months after stroke.(Zigmond & Snaith, 1983) A STROBE flow chart demonstrates the cohort follow up. (Figure 1)

Assessment methods and procedures:

The World Health Organization (WHO) definition of stroke was used. (Hatano, 1976) To increase the completeness of notification sixteen overlapping referral sources (primary care and hospitals for example) were used and case ascertainment was estimated as 88% using a capture-recapture model. (Heuschmann et al., 2008; Tilling, Sterne, & Wolfe, 2001) Data collected during the acute phase of stroke included socio-demographic factors, medication used before and after stroke, comorbidities and stroke severity, including Glasgow coma scale (Jennett & Bond, 1975) categorized as severe impairment, score (3–8), moderate (9–12), and mild (13–15), incontinence, and paresis. This information is collected from medical notes by the SLSR field workers. A senior stroke physician verifies the clinical data of all patients being registered. Patients were assessed at 3 months after stroke, at one year, then annually. Follow up at 3 months after stroke was by postal questionnaire or interview. At follow up patients were assessed for disability using the Barthel Index (BI) (Wolfe, Taub,

Woodrow, & Burney, 1991), categorised as severe disability (0-14), moderate (15-19) and independent (20).

HADS was used to screen for anxiety and depression. Since HADS cannot be answered by proxy, all information was collected directly from patients. While patients with some degree of cognitive or communication impairment can respond to HADS, no data could be collected from patients with severe cognitive or communication impairment, from field worker, or the patient's next of kin in case of postal questionnaire, judged would give invalid responses.

The scale comprises 14 items, 7 items screen for anxiety, 7 for depression, and all 14 items assess emotional distress. Each item takes four possible response options [0-3] making the possible sum scores range from 0 to 21 for each of the subscales, HADS-A and HADS-D. Items are coded in one direction such that higher values indicate greater symptom severity. A score ≥ 8, has been proposed for the identification of caseness, for both depression and anxiety in patients with different physical health conditions, with a sensitivity of round 0.8 (Bjelland et al., 2002).

Statistical Analysis:

To test the uni-dimensionality assumption required for the application of IRT models, factor analysis was performed. Both the maximum likelihood (ML) and principal component factor (PCF) analysis methods were used to extract factors. The former method is known to be more robust to violations of the normality assumptions. (Acock, 2013) Confirmatory factor analysis (CFA) was used, with two factors for the full scale and a single factor applied separately to each sub-scale.

The two-parameter logistic (2PL) IRT model was initially used to estimate the parameters, difficulty and discrimination. Where items were treated as binary, the ordinal original categories [0-3] were transformed into binary (0 / 1), "0" representing scores of 0 and 1, and "1" representing scores of 2 and 3. The properties of the items were displayed graphically using, Item Characteristic Curves (ICCs): a mathematical expression that relates a probability of choice of a response category, in an item to the trait being assessed by the scale. Item information functions (IIFs) were also derived and presented for items with highest and lowest discrimination for each subscale. The height of an IIF represents the amount of information an item provides about the difficulty parameter, and is proportional to the item's estimated discrimination. While the properties of all the items were summarised in tables, we have given more attention and/or graphically presented items that have special characteristics, for example items with best discrimination, poor discrimination and items that have similar properties. Such items would be of potential use for the modification of the scale.

The difficulty parameter (location) β, can be interpreted as an indicator of how likely a patient is to choose a certain level of response (category) in an item, based on the underlying level of trait (anxiety / depression). Low levels of a trait would be associated with lower categories of response, for an item.

Item discrimination α, indicates how well item separates patients with different underlying trait (anxiety / depression). An item with good discrimination would have a large difference in the probability of positive response (choosing "1" in a binary item to mean limited/ or limited a lot as opposed to choosing "0" which means not limited at all, or little limitations for example) above and below its location, whereas for items with low discrimination, the difference in probabilities would be small and gradual. The steepness of the item characteristic curve (ICC) in its middle section, estimates the item's discrimination. Steeper curves are better at separating respondents. Items with large magnitude have strong

associations with the underlying trait, and vice versa. Guidelines for grading discrimination have interpreted estimates below 0.65 as low; 0.65-1.34 moderate; 1.35-1.69 high, and 1.7 or more very high. (Baker, 1985). Graded response models (GRM) were used to detect further details on responses to scale items in their ordinal form. The statistical software Stata (14.0) was used for the analysis. (StataCorp, 2015)

Results

The mean age was 68.07 years (SD=14.2) and 54.7% of the patients were males. The baseline demography and stroke severity of the included patients are described in Table 1. Values are given for the full cohort and for patients fulfilling the criteria for HADS-D caseness. No significant differences were observed when classification was based on the same threshold for HADS-A. (results not presented)

Testing the dimensionality of HADS

A two-factor solution fits the data adequately where the full scale (14 items) was examined. Supplementary Table 1, (a) demonstrates that all items loaded positively towards the first factor while most items from HADS-A loaded negatively towards the second factor. Eigen values, were 5.76, and 1.37 for the first and second factors respectively, using the (PCF) and 5.21, and 0.88 using the (ML) approach, Supplementary Table 1, (b). Despite an Eigen value of less than 1, for the ML method, thus failing to fulfil the conventional criteria for inclusion in the final solution (Rietveld & Van Hout, 1993), there were no considerable differences in the signs or magnitudes of

loadings, for the factors extracted by the two methods.

When CFA was applied to each domain, first specifying two factors, then a single factor, the single factor model provided a better fit for each, whether ML or PCF analysis was used. All items loaded positively and significantly towards the first factor.

The goodness of fit for each was 'very good', as assessed by the small standardized root mean squared residual (SRMSR) ≤ 0.05, and a high comparative fit index (CFI) ≥ 0.90 criteria. (Acock, 2013), Supplementary Table S1, (c).

Supplementary Table S1 (d), summarises the estimated correlations between items within each subscale of HADS and the underlying factor it measures. The item test correlation, was reasonably similar across all items within subscales as expected, except for "Q5", from HADS-D, which has lower correlation with the scale. Item-rest correlation that indicates the correlation of the item with the rest of the subscale its part of, have shown moderate correlation for all items, except Q8 form HAD-A (correlation coefficient = 0.49) and Q5 from HADS-D (correlation coefficient = 0.42), that have shown relatively lower correlations. Cronbach's alpha that measures the internal consistency was generally acceptable but not high for the two subscales, and slightly higher for HADS-A items, the overall estimate was 0.84 and 0.81 for HADS-A and HADS-D respectively.

The psychometric properties of HADS

Estimates based on the 2PL model show low discrimination for Q5 and Q8, and the highest discrimination values were found for Q4 and Q13. Table (2)

The ICCs, for items with the highest and lowest discrimination are displayed in (Figure 2), for the two domains of HADS. The vertical axis, represents the probability of positive response (choice that indicates more symptoms), and the horizontal axis represents the trait level (θ), with average, θ = 0. For items with high discrimination (Q4 and Q13, solid lines) in each sub-scale, the probability of a positive response increases sharply and more rapidly as the latent trait (Depression /Anxiety) increases above the difficulty level, showing higher capability of differentiating between patients with different levels of symptoms, whereas for items (Q5 and Q8) with low discrimination (Dashed line) the increase is more gradual.

For Q5 and Q4, presented in panel (a), the patient must have a depression level of approximately -.068, and 1.55 (in trait metric), to have a 0.5 probability of positive response for the two items respectively. Panel (b) presents the corresponding values of 0.87 and 1.15, for Q8 and Q13 respectively, from HADS-A.

The item information curves (IICs) for the 4 items are displayed in panels (a) and (b), Figure 3. More information was provided by items with high discrimination (Q4 and Q13), with peak values where the underlying construct was positive. Items with lowest discrimination provided poor information over a wider range of the underlying construct -2 to 2, and -1 to 2, for Q5 and Q8, respectively.

Considering the properties of items over time, the ranking of discrimination and difficulty at each assessment point was maintained throughout the follow up period from 3 months to 5 years after stroke, with few exceptions for which the difference in magnitude was negligible. (Supplementary Figure 1) The figure, also demonstrates that the items, Q3, Q6, and Q9 have similar discrimination, therefore may convey similar information.

The GRM results agreed well with those based on the 2PL model. As the GRM is an ordered logistic model, difficulty parameters are naturally estimated in an increasing order for each item. Looking at the middle item in order of discrimination for example, Q2 from HADS-D, (Table 3): a person with an underlying level of depression, $\theta = -0.49$, has a 50% chance of answering "0" rather than "≥ 1"; a person with $\theta = 0.74$ has a 50% chance of answering "0" or "1" rather than "≥2"; and a person with $\theta = 1.39$ has a 50% chance of answering 0, 1, or 2 rather than "3".

Supplementary Figure 2, presents the frequencies of response to each of the categories (0-3) for items with lowest and highest discrimination from HADS-D and HADS-A, namely, Q5, Q4, Q8 and Q13.

Figure 4, presents category characteristic curves that illustrates the probability of response to each category in relation to the (latent trait) underlying level of symptoms. For Q8, for example, respondents with the latent trait value θ, below approximately -0.5 were most likely to respond in the first category; those with θ between -0.5 to 1.0, were most likely to respond in the second category; those with θ approximately 1.0 to 2.0, were most likely to respond in the third category and those with $\theta \geq 2.0$, were most likely to respond in the fourth category. The other three items differ in how response to each item relates to the underlying trait. Items with similar discrimination have similar features of response.

Discussion

This study on stroke patients, supplements the evidence on the presence of two dimensions for HADS, and a single dimension for each of its subscales. HADS dimensionality has been examined previously and inconsistencies noted with systematic reviews reporting between one to four underlying constructs, although the majority reported two domains representing anxiety and depression.(Cosco, Doyle, Ward, & McGee, 2012; Johnston, Pollard, & Hennessey, 2000) Our findings favouring two domains for HADS in stroke patients, are consistent with others conducted on the general population, somatic patients and psychiatric patients. (Bjelland et al., 2002; Matsudaira et al., 2009; Zigmond & Snaith, 1983) Internal consistency, as assessed by Cronbach's alpha coefficient, was 0.84 and 0.81 for HADS-A and HADS-D respectively, agreeing with similar estimates reported for the subscales in a systematic review including 15 studies. (Bjelland et al., 2002)

The study identified items with varying psychometric characteristics, including varying difficulty and discrimination. High discrimination items offer clinician greater efficiency in determining caseness. From the depression domain (HADS-D), the item, ‘I can laugh and see

the funny side of things' (Q4), consistently showed the highest discrimination, whereas, 'I feel as if I am slowed down' (Q5), consistently showed the lowest. Previous reports have also identified Q5 as poorly performing, noting that it may measure somatic rather than emotional symptoms, counter to the design intentions of HADS-D. (Johnston et al., 2000; Snaith & Zigmond, 1986) From the anxiety domain (HADS-A), the item, 'I get sudden feelings of panic' (Q13), demonstrated the highest discrimination, whereas the item, 'I feel restless as if I have to be on the move' (Q8) had the lowest discrimination.

Studies that used IRT methods to investigate the performance of HADS in other patients' groups have mostly used the one parameter (1P) Rasch model, and did not attempt to use the 2PL model, or GRM, making comparisons with our findings not straightforward in most situations. In Motor Neurone disease (MND), for example, the model did not fit well, Q5 was thought to be confounded with somatic symptoms, Q3 and Q9, have shown dependency, and the study suggested the exclusion of 3 items from HADS total. (Gibbons et al., 2011) For patients with epilepsy, similarly Rasch model was used, the scale was shown to have high internal consistency, and to perform equally well in male and female and across two types of epilepsy. (Forjaz, Rodriguez-Blazquez, Martinez-Martin, & Longitudinal Parkinson's Disease Patient Study, 2009) For patients attending an out-patient musculoskeletal rehabilitation program, the model fit was generally poor, and the item "can sit at ease and feel relaxed" from HADS-A was highlighted as a possible redundant item. (Pallant & Tennant, 2007)

We have used the 2PL model, in addition to the GRM, aiming to provide a more complete picture on the properties of HADS, based on the original ordinal type of items, as well as in a simplified binary form that would allow the comparison of HADS with other measures that are often used with binary response items, such as the General Health Questionnaire (GHQ), in detecting psychiatric morbidity in stroke patients. (O'Rourke, MacHale, Signorini, & Dennis, 1998).

Our study provided first time details about the properties of HADS items based on 2PL models, and GRM models, in a large sample of stroke patients. Items that provide little information about the underlying traits had relatively large proportions of patients reporting higher levels of symptoms. For example, for (Q5), 15% of patients selected the lowest level of "not at all", while the other three categories were chosen by 33%, 21% and 31%, respectively. The fact that 52% of patients selected the highest levels of symptoms, suggests that the item is insufficiently precise at identifying depression severity, as also indicated by its poor discrimination. The corresponding proportions for a difficult item with high discrimination (Q4) were 69% for the lowest level of symptoms -first category of the item, followed by 21%, 7% and 3% for the subsequent three higher levels of symptoms, respectively. Since only 10% of patients selected either of the two categories indicating severe symptoms, this suggest that the item is difficult, and is useful in identifying patients with more serious symptoms. Such items, may be utilised in the development of computerized adaptive testing (CAT) techniques that administer items tailored to the level of the underlying trait. (Chakravarty et al., 2007; Ware, 2003) Difficult or medium difficulty items, may be used as a starting point for assessment, followed by other items with an appropriate level for the respondent.(Hays et al., 2000) Such an approach has the potential to decrease the burden of questions on patients while increasing the efficiency of the assessment and the precision of the measure.

Previous studies have shown considerable variations in the recommended cut-off points for caseness determination, and have revealed that these were suboptimal. (Aben et al., 2002; First & Pincus, 2002; Sagen et al., 2009; Turner et al., 2012) The variations in the properties of items shown by the current study, suggest that giving equal weight to all items, to determine cut-off points for caseness, based on summation scores, may be inappropriate.

Within patients' groups (stroke patients for example), giving equal weights may be reasonable; assuming patients have similar priorities, would be equally affected by weights, and comparisons based on caseness determined by cut-off points, would therefore be valid. These weighting assumptions however, are unlikely to generalise to populations with different physical conditions, since physical debility is likely to differ according to the spectrum of physical problems associated with each illness.

While our study, is a register based and we don't have a gold standard assessment such as the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) to formally compare a range of thresholds with, or to develop subset of items using the properties of HADS, such progress may be achieved in future studies designed to use DSM-IV, alongside HADS. In addition, clinical trials may serve to test the performance of modified versions of the scale, and to determine whether shorter versions would be superior to the standard form in detecting psychiatric symptoms and in any other features such as achieving better patient satisfaction.

The study has also shown that HADS-A contains three items that convey similar information: (Q3), 'I get a sort of frightened feeling as if something awful is about to happen'; (Q6), 'I get a sort of frightened feeling like butterflies in my stomach'; and (Q9) 'Worrying thoughts go through my mind'. These three items have close discrimination estimates (2.28, 2.58 and 2.41, respectively) and their meaning appears to overlap. Clinical trials are often constrained in sample size and budget, may require fewer subjects, and greater assurances may be given that the perspectives of the patients are included. Further studies to investigate how stroke patients perceive these three questions may provide required information to confirm whether any of the questions could be considered redundant.

Strengths: This study represents the first use of IRT methods in a large sample of stroke patients to identify item properties in HADS. The sample including 1443 patients, was taken

over a long period of up to 5 years of follow up, post stroke. HADS completion rates in this sample with physical comorbidity was relatively high at over 75%. The study conformed to the recommendations of the STROBE statement.(Gallo et al., 2012)

Limitations*:* approximately 20% of patients were lost to follow up; this may have introduced some bias and women had slightly higher rates of loss to follow up. Cognitive or communication difficulties contributed to non-response. Nevertheless, estimates from analyses of patients with complete data showed no significant differences in the baseline characteristics, compared with those lost to follow-up. The consistency of the properties reported in this study, would benefit from external validation using large external samples of stroke patients and patients with other physical morbidities.

Acknowledgements

This research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health."

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Authors statement:

Authors' contributions: All authors have contributed to the intellectual content, interpretation of data, and have critically reviewed the paper at all stages. Salma Ayis did the statistical analysis, and drafted the first version of the paper. All authors contributed to the writing and have approved the final copy.

Conflicts of interest

None of the authors: Salma Ayis, Luis Ayerbe, Mark Ashworth, and Charles Wolfe has any conflicts of interest to disclose.

Funding

The study was funded by Guy's and St Thomas' Hospital Charity, The Stroke Association, Department of Health HQIP Grant, UK, National Institute for Health Research Programme Grant (RP-PG-0407-10184). The authors (CW) acknowledge financial support from the Department of Health via the National Institute for Health Research (NIHR) Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. CDAW is an NIHR Senior Investigator. This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10184). The fundes have no role in this research.

Ethics

Patients or their relatives gave written informed consent to the SLSR field workers at the time of registration. The ethics committees of Guy's and St. Thomas' Hospital National Health Service Foundation Trust, King's College Hospital Foundation, National Hospital for Nervous Diseases, Queen's Square Hospital, St. George's Hospital, and Westminster Hospital approved the study.

References

Aben, I., Verhey, F., Lousberg, R., Lodder, J., & Honig, A. (2002). Validity of the beck depression inventory, hospital anxiety and depression scale, SCL-90, and hamilton depression rating scale as screening instruments for depression in stroke patients. *Psychosomatics, 43*(5), 386-393. doi:10.1176/appi.psy.43.5.386

Acock, A. C. (2013). *Discovering Structural Equation Modeling Using Stata, Revised Edition*: Stata Press.

Andersson, E. (1993). The Hospital Anxiety and Depression Scale: homogeneity of the subscales. *Social Behavior and Personality: an international journal, 21*(3), 197-204.

Ayerbe, L., Ayis, S., Wolfe, C. D., & Rudd, A. G. (2013). Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *Br J Psychiatry, 202*(1), 14-21. doi:10.1192/bjp.bp.111.107664

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry, 161*(12), 2163-2177. doi:10.1176/appi.ajp.161.12.2163

Baker, F. B. (1985). *The basics of item response theory*. Portsmouth, N.H.: Heinemann.

Barker-Collo, S. L. (2007). Depression and anxiety 3 months post stroke: prevalence and correlates. *Arch Clin Neuropsychol, 22*(4), 519-531. doi:10.1016/j.acn.2007.03.002

Barth, J., & Martin, C. R. (2005). Factor structure of the Hospital Anxiety and Depression Scale (HADS) in German coronary heart disease patients. *Health and Quality of Life Outcomes, 3*(1), 1.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res, 52*(2), 69-77.

Caci, H., Baylé, F. J., Mattei, V., Dossios, C., Robert, P., & Boyer, P. (2003). How does the Hospital and Anxiety and Depression Scale measure anxiety and depression in healthy subjects? *Psychiatry research, 118*(1), 89-99.

Campbell Burton, C. A., Murray, J., Holmes, J., Astin, F., Greenwood, D., & Knapp, P. (2013). Frequency of anxiety after stroke: a systematic review and meta-analysis of observational studies. *Int J Stroke, 8*(7), 545-559. doi:10.1111/j.1747-4949.2012.00906.x

Chakravarty, E. F., Bjorner, J. B., & Fries, J. F. (2007). Improving patient reported outcomes using item response theory and computerized adaptive testing. *J Rheumatol, 34*(6), 1426-1431.

Cosco, T. D., Doyle, F., Ward, M., & McGee, H. (2012). Latent structure of the Hospital Anxiety And Depression Scale: a 10-year systematic review. *J Psychosom Res, 72*(3), 180-184. doi:10.1016/j.jpsychores.2011.06.008

Dawkins, N., Cloherty, M. E., Gracey, F., & Evans, J. J. (2006). The factor structure of the Hospital Anxiety and Depression Scale in acquired brain injury. *Brain injury, 20*(12), 1235-1239.

First, M. B., & Pincus, H. A. (2002). The DSM-IV Text Revision: rationale and potential impact on clinical practice. *Psychiatr Serv, 53*(3), 288-292. doi:10.1176/appi.ps.53.3.288

Forjaz, M. J., Rodriguez-Blazquez, C., Martinez-Martin, P., & Longitudinal Parkinson's Disease Patient Study, G. (2009). Rasch analysis of the hospital anxiety and depression scale in Parkinson's disease. *Mov Disord, 24*(4), 526-532. doi:10.1002/mds.22409

Gallo, V., Egger, M., McCormack, V., Farmer, P. B., Ioannidis, J. P., Kirsch-Volders, M., . . . Vineis, P. (2012). STrengthening the Reporting of OBservational studies in Epidemiology--Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. *Mutagenesis, 27*(1), 17-29. doi:10.1093/mutage/ger039

Gibbons, C. J., Mills, R. J., Thornton, E. W., Ealing, J., Mitchell, J. D., Shaw, P. J., . . . Young, C. A. (2011). Rasch analysis of the hospital anxiety and depression scale (HADS) for use in motor neurone disease. *Health Qual Life Outcomes, 9*, 82. doi:10.1186/1477-7525-9-82

Hackett, M. L., & Anderson, C. S. (2005). Predictors of depression after stroke: a systematic review of observational studies. *Stroke, 36*(10), 2296-2301. doi:10.1161/01.STR.0000183622.75135.a4

Hackett, M. L., & Pickles, K. (2014). Part I: frequency of depression after stroke: an updated systematic review and meta-analysis of observational studies. *Int J Stroke*. doi:10.1111/ijs.12357

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park ; London: Sage.

Hatano, S. (1976). Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ, 54*(5), 541-553.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Med Care, 38*(9 Suppl), II28.

Heuschmann, P. U., Grieve, A. P., Toschke, A. M., Rudd, A. G., & Wolfe, C. D. (2008). Ethnic group disparities in 10-year trends in stroke incidence and vascular risk factors: the South London Stroke Register (SLSR). *Stroke, 39*(8), 2204-2210. doi:10.1161/STROKEAHA.107.507285

Jennett, B., & Bond, M. (1975). Assessment of outcome after severe brain damage. *Lancet, 1*(7905), 480-484.

Johnston, M., Pollard, B., & Hennessey, P. (2000). Construct validation of the hospital anxiety and depression scale with clinical populations. *J Psychosom Res, 48*(6), 579-584.

Leppavuori, A., Pohjasvaara, T., Vataja, R., Kaste, M., & Erkinjuntti, T. (2003). Generalized anxiety disorders three to four months after ischemic stroke. *Cerebrovasc Dis, 16*(3), 257-264. doi:71125

Matsudaira, T., Igarashi, H., Kikuchi, H., Kano, R., Mitoma, H., Ohuchi, K., & Kitamura, T. (2009). Factor structure of the Hospital Anxiety and Depression Scale in Japanese psychiatric

outpatient and student populations. *Health Qual Life Outcomes, 7*, 42. doi:10.1186/1477-7525-7-42

Mykletun, A., Stordal, E., & Dahl, A. A. (2001). Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *The British journal of psychiatry, 179*(6), 540-544.

Naess, H., Waje-Andreassen, U., Thomassen, L., Nyland, H., & Myhr, K. M. (2006). Health-related quality of life among young adults with ischemic stroke on long-term follow-up. *Stroke, 37*(5), 1232-1236. doi:10.1161/01.STR.0000217652.42273.02

O'Rourke, S., MacHale, S., Signorini, D., & Dennis, M. (1998). Detecting psychiatric morbidity after stroke: comparison of the GHQ and the HAD Scale. *Stroke, 29*(5), 980-985.

Olino, T. M., Yu, L., McMakin, D. L., Forbes, E. E., Seeley, J. R., Lewinsohn, P. M., & Pilkonis, P. A. (2013). Comparisons across depression assessment instruments in adolescence and young adulthood: an item response theory study using two linking methods. *J Abnorm Child Psychol, 41*(8), 1267-1277. doi:10.1007/s10802-013-9756-6

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol, 46*(Pt 1), 1-18.

Razavi, D., Delvaux, N., Farvacques, C., & Robaye, E. (1990). Screening for adjustment disorders and major depressive disorders in cancer in-patients. *The British Journal of Psychiatry, 156*(1), 79-83.

Rietveld, T., & Van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*: Walter de Gruyter.

Sagen, U., Vik, T. G., Moum, T., Mørland, T., Finset, A., & Dammen, T. (2009). Screening for anxiety and depression after stroke: Comparison of the Hospital Anxiety and Depression Scale and the Montgomery and Åsberg Depression Rating Scale. *J Psychosom Res, 67*(4), 325-332.

Snaith, R. P. (2003). The Hospital Anxiety And Depression Scale. *Health Qual Life Outcomes, 1*, 29. doi:10.1186/1477-7525-1-29

Snaith, R. P., & Zigmond, A. S. (1986). The hospital anxiety and depression scale. *Br Med J (Clin Res Ed), 292*(6516), 344.

StataCorp. (2015). Stata Statistical Software: . *Release 14. College Station, TX: StataCorp LP*.

Terol-Cantero, M. C., Cabrera-Perona, V., & Martín-Aragón, M. (2015). Hospital Anxiety and Depression Scale (HADS) review in Spanish Samples. *Anales de Psicología/Annals of Psychology, 31*(2), 494-503.

Tilling, K., Sterne, J. A., & Wolfe, C. D. (2001). Estimation of the incidence of stroke using a capture-recapture model including covariates. *Int J Epidemiol, 30*(6), 1351-1359; discussion 1359-1360.

Townend, B. S., Whyte, S., Desborough, T., Crimmins, D., Markus, R., Levi, C., & Sturm, J. W. (2007). Longitudinal prevalence and determinants of early mood disorder post-stroke. *J Clin Neurosci, 14*(5), 429-434. doi:10.1016/j.jocn.2006.01.025

Turner, A., Hambridge, J., White, J., Carter, G., Clover, K., Nelson, L., & Hackett, M. (2012). Depression Screening in Stroke A Comparison of Alternative Measures With the Structured Diagnostic Interview for the Diagnostic and Statistical Manual of Mental Disorders, (Major Depressive Episode) as Criterion Standard. *Stroke, 43*(4), 1000-1005.

Ware, J. E. (2003). Conceptualization and measurement of health-related quality of life: comments on an evolving field. *Arch Phys Med Rehabil, 84*, S43-S51.

Wolfe, C. D., Taub, N. A., Woodrow, E. J., & Burney, P. G. (1991). Assessment of scales of disability and handicap for stroke patients. *Stroke, 22*(10), 1242-1244.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatr Scand, 67*(6), 361-370.

TABLE 1. Demography and stroke severity in the acute phase

Depression score	**HADS-D (<8)**		**HADS-D (8+)**		**all**
Number of patients (n)	986		457		1443
Age (SD)	68.38 (14.11)		67.33		68.05
Gender					
Male	557	56.49%	233	50.98%	54.75%
Female	429	43.51%	224	49.02%	45.25%
Barthel Index (acute Phase)					
Severe (0-14)	273	27.69%	200	43.76%	32.78%
Mild (15-19)	179	18.15%	97	21.23%	19.13%
Independent (20)	382	38.74%	95	20.79%	33.06%
Unknown	152	15.42%	65	14.22%	15.04%
Lift arm					
Unable	142	14.40%	86	18.82%	15.80%
Able	455	46.15%	158	34.57%	42.48%
Unknown	389	39.45%	213	46.61%	41.72%
Can walk					
Unable	238	24.14%	111	24.29%	24.19%
Able	358	36.31%	131	28.67%	33.89%
Unknown	390	39.65%	215	47.04%	41.92%
GC Scale					
Severe impairment (3-8)	16	1.74%	16	3.74%	2.38%
Moderate (9-12)	71	7.74%	35	8.18%	7.88%
Mild (13-15)	830	90.51%	377	88.08%	89.74%
Incontinence					
No	769	77.99%	301	65.86%	74.15%
Yes	187	18.97%	134	29.32%	22.25%
Unknown	30	3.04%	22	4.82%	3.60%
Paresis					
No	182	18.46%	58	12.69%	16.63%
Yes	425	43.10%	231	50.55%	45.46%
Unknown	379	38.44%	168	36.76%	37.91%

Note: HADS-D: Hospital Anxiety and Depression Scale (HADS), depression domain; HADS-A: Hospital Anxiety and Depression Scale (HADS), anxiety domain; GC: Glasgow Coma scale; SD: standard deviation.

Table 2. Difficulty and Discrimination for HADS-A and HADS-D items, 3 Months after Stroke, Using 2 Parameter Logistic (2PL) Models and Binary Response Forms

	Discrimination	**Difficulty**

Question	Depression (HADS-D)	Estimate	SE	Estimate	SE
Q5	I feel as if I am slowed down	1.23	0.11	-0.07	0.06
Q7	I have lost interest in my appearance	1.39	0.13	1.46	0.10
Q14	I can enjoy a good book, radio or TV programme	1.49	0.16	1.91	0.14
Q2	I still enjoy the things I used to enjoy	1.99	0.18	0.68	0.05
Q12	I look forward with enjoyment to things	2.11	0.19	1.04	0.06
Q10	I feel cheerful	2.29	0.22	1.26	0.07
Q4	I can laugh and see the funny side of things	2.44	0.26	1.55	0.08
	Anxiety (HADS-A)				
Q8	I feel restless as if I have to be on the move	1.35	0.11	0.87	0.07
Q11	I can sit at ease and feel relaxed	1.71	0.15	1.30	0.08
Q1	I feel tense or wound up	1.74	0.15	1.06	0.07
Q3	I get a sort of frightened feeling as if something awful is about to happen	2.28	0.20	0.89	0.05
Q9	Worrying thoughts go through my mind	2.41	0.21	0.78	0.05
Q6	I get a sort of frightened feeling like butterflies in my stomach	2.58	0.25	1.37	0.07
Q13	I get sudden feelings of panic	3.34	0.35	1.15	0.05

Note: SE = standard error

Table 3. Difficulty and Discrimination for HADS-A and HADS-D items, 3 Months after Stroke, Using Graded Response Models (GRMs) and the Original Ordinal Response Categories

Depression (HADS-D)				**Anxiety (HADS-A)**			
Que	**Discrimination:**	**Esti**	**S**	**Que**		**Esti**	**S**
Q5	I feel as if I am slowed down	1.09	0.07	Q8	I feel restless as if I have to be on the move	1.27	0.08
Q7	I have lost interest in my appearance	1.50	0.10	Q11	I can sit at ease and feel relaxed	1.59	0.09
Q14	I can enjoy a good book, radio or TV programme	1.52	0.10	Q1	I feel tense or wound up	1.60	0.09
Q2	I still enjoy the things I used to enjoy	1.87	0.11	Q9	Worrying thoughts go through my mind	1.94	0.11
Q12	I look forward with enjoyment to things	2.21	0.13	Q6	I get a sort of frightened feeling like butterflies in my stomach	2.35	0.14
Q10	I feel cheerful	2.34	0.14	Q3	I get a sort of frightened feeling as if something awful is about to happen	2.42	0.14

Q4	I can laugh and see the funny side of things	2.64	0.18	Q13	I get sudden feelings of panic	2.82	0.17

Depression (HADS-D)			**Anxiety (HADS-A)**		
Difficulty:	**Esti**	**SE**		**Esti**	**SE**
I feel as if I am slowed down ≥ 1	-1.92	0.12	I feel restless as if I have to be on the	-0.47	0.06
I feel as if I am slowed down ≥ 2	-0.05	0.06	I feel restless as if I have to be on the	0.93	0.07
I feel as if I am slowed down ≥ 3	0.94	0.08	I feel restless as if I have to be on the	2.26	0.13
I have lost interest in my	0.49	0.05	I can sit at ease and feel relaxed ≥ 1	-0.12	0.05
I have lost interest in my	1.41	0.08	I can sit at ease and feel relaxed ≥ 2	1.37	0.07
I have lost interest in my	2.21	0.12	I can sit at ease and feel relaxed ≥ 3	2.59	0.13
I can enjoy a good book, radio or	0.64	0.05	I feel tense or 'wound up ≥ 1	-0.61	0.05
I can enjoy a good book, radio or	1.87	0.10	I feel tense or 'wound up ≥ 2	1.11	0.06
I can enjoy a good book, radio or	2.62	0.15	I feel tense or 'wound up ≥ 3	1.98	0.10
I still enjoy the things I used to	-0.49	0.05	Worrying thoughts go through my mind	-0.27	0.05
I still enjoy the things I used to	0.74	0.05	Worrying thoughts go through my mind	0.86	0.05
I still enjoy the things I used to	1.39	0.07	Worrying thoughts go through my mind	1.65	0.08
I look forward with enjoyment to	0.08	0.04	I get a sort of frightened feeling like butterflies in my stomach ≥ 1	0.35	0.04
I look forward with enjoyment to	1.04	0.05	I get a sort of frightened feeling like butterflies in my stomach ≥ 2	1.41	0.06
I look forward with enjoyment to	1.74	0.08	I get a sort of frightened feeling like butterflies in my stomach ≥ 3	2.16	0.10
I feel cheerful ≥ 1	-0.08	0.04	I get a sort of frightened feeling as if something awful is about to happen ≥ 1	0.21	0.04
I feel cheerful ≥ 2	1.26	0.06	I get a sort of frightened feeling as if something awful is about to happen ≥ 2	0.87	0.05
I feel cheerful ≥ 3	2.00	0.09	I get a sort of frightened feeling as if	1.73	0.08
I can laugh and see the funny side	0.61	0.04	I get sudden feelings of panic ≥ 1	0.15	0.04
I can laugh and see the funny side	1.50	0.07	I get sudden feelings of panic ≥ 2	1.20	0.05
I can laugh and see the funny side	2.23	0.10	I get sudden feelings of panic ≥ 3	2.12	0.09

Note: SE = standard error; ≥ 1, ≥ 2, ≥ 3: stands for the difficulty of choosing category above or equal to the cut off points 1, 2 and 3 respectively. Items codes for the Graded Response Model (GRM) were 1, 2, 3 and 4.

Figures Legend. HADS-D, items: Q4 “I can Laugh and see the funny side of things”, Q5 “I feel as if I am slowed down". HADS-A, items: Q13 “I get sudden feelings of panic”, Q8 “I feel restless as if I have to be on the move"

Figure 1.

Figure 1. Flow chart showing the number of stroke patients included at each follow-up

Figure 1 Legend: N= number of patients interviewed; LTF=Lost to follow up; FU= completed the follow up; HAD: HADS completed.

Figure 2.

Item Characteristic Curves for Items with Highest and Lowest Discrimination of the Sub-scales HADS-D and HADS-A

Figure 3.

Item Information Functions for Items with Highest and Lowest Discrimination of the Sub-scales HADS-D and HADS-A

Figure 4.

Category Characteristics Curves for Items with Highest and Lowest Discrimination of the Sub-scales HADS-D and HADS-A

Figures (2-4) Legend. HADS: Hospital Anxiety and Depression Scale.
HADS-D Items: Q4 “I can Laugh and see the funny side of things”, Q5 “I feel as if I am slowed down".
HADS-A items: Q13 “I get sudden feelings of panic”, Q8 “I feel restless as if I have to be on the move".